

# Scalar vs. mereological conceptualizations of the N-BY-N and NUM-BY-NUM adverbials

Lucia Vlášková  
Masaryk University  
Brno, Czech Republic  
l.vlaskova@gmail.com

Mojmír Dočekal  
Masaryk University  
Brno, Czech Republic  
docekal@phil.muni.cz

## Abstract

The multiword adverbials N-BY-N and NUM-BY-NUM (as English *brick by brick* and *one by one*, respectively) are event modifiers which require temporal sequencing of the event they modify into a linearly ordered series of sub-events. Previous studies unified these two constructions under a single semantic analysis and adopted either a mereological or a scalar approach. However, based on a corpus study examining new Slavic language material and a binomial logistic regression modelling of the manually annotated data, we argue that two separate analyses are needed to account for these constructions, namely a scalar analysis for the N-BY-N construction and a mereological one for the NUM-BY-NUM construction.

## 1 Background

Sentences in (1) both contain constructions traditionally analyzed as adverbials consisting of multiple words. In (1-a), it is a so-called N-BY-N construction, where two identical nouns are connected by a preposition; in (1-b), a so-called NUM-BY-NUM construction, where a preposition connects two numerals. It is these multiword adverbials, their semantic analysis and a new formal description that will be the focus of this paper.

- (1) a. The workers built the house *brick by brick*. N-BY-N  
b. They laid the bricks *one by one*. NUM-BY-NUM

### 1.1 Previous analyses

Formal semantic descriptions of both N-BY-N and NUM-BY-NUM constructions are of the two general types: a mereological and a scalar one.

The mereological approach assumes a division of the main event into a sequence of temporally ordered sub-events, either with the help of a particular syntactic structure at the LF level (Beck and von Stechow, 2007) or by utilizing relevant  $\theta$ -roles (Brasoveanu and Henderson, 2009). In result, both events and participating pluralities are simultaneously divided and linked together in a distributive manner.

On the other hand, the scalar approach developed mainly by Henderson (2013) and also Braginsky and Rothstein (2008) formalizes the contribution of the adverbial as a fixation of a scalar change. More precisely, the adverbials in question require the event to be divided into a plurality of temporally sequenced sub-events, just like the mereological accounts, but the sub-events here are distinguished by the scalar interval that sets the unit of the scalar change (the N in N-BY-N) that the theme of the main verb undergoes, without necessarily dividing the theme itself. Furthermore, Henderson (2013) ties the N-BY-N modification closely to the verbs of scalar change, that is, incremental theme verbs (*John ate the pie*), change of state verbs (*The crack widened*) and inherently directed motion verbs (*John walked to the store*), arguing for their unified scalar semantics based on the accessible N-BY-N modification.

Although sometimes differences between the N-BY-N and NUM-BY-NUM construction are acknowledged, the usual tacit assumption of both mereological and scalar approaches is that one framework is

---

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

expressive enough to describe both types of constructions. We challenge this assumption and claim that a realistic description of the N-BY-N and the NUM-BY-NUM constructions must be both mereological *and* scalar.

## 1.2 New data

Our empirical arguments for a new analysis come from a corpus query, manual annotation, and binomial logistic regression modelling of the tagged data.

We worked with Czech, a West Slavic language with around 13 million native speakers and rich morphology that reveals some properties obscured in English. Let us look more closely on the Czech equivalent of the previous example in (2).

- (2) a. Dělníci postavili dům cihlu po cihle.  
workers built.PL house.MASC.A brick.FEM.A after brick.FEM.LOC  
'The workers built the house brick by brick.' N-BY-N
- b. Ukládali cihly jednu za druhou.  
laid.3PL bricks.A one.FEM.A after second.FEM.INST  
'They laid the bricks one by one.' NUM-BY-NUM

Firstly, it is important to note that there are two alternative realizations of the prepositions in either N-BY-N or NUM-BY-NUM construction: the preposition *po* "after", seen in (2-a), and the preposition *za* "after", seen in (2-b), which will be the focus of our research, although both variants behave similarly with respect to their frequency and collocations. The existence of alternative prepositions is recognized also in English where the possible candidates are: *by*, *after*, *upon*, *over*, *within* (Beck and von Stechow, 2007).

Due to relatively transparent morphology, we see that the case on the nouns and numerals is assigned by the respective preposition. Moreover, the gender agreement of the relevant expressions (nouns/numerals) is active only in the NUM-BY-NUM construction where it is guided by the corresponding noun (*bricks* in (2-b)) and applies to both numerals, hence a mix of two genders within a single construction is not grammatical.

The most noticeable distinctive property in (2-b) is the usage of different lemmas within the construction, i.e. *jeden* "one" and *druhý* "second", as opposed to English *one* repeated in both slots. In this respect, the Czech construction is similar to English *one after the other*, which is considered to be semantically closely related to the NUM-BY-NUM construction (Beck and von Stechow, 2007), but contrary to English, the Czech adverbial with the same lemma, *\*jeden za jedním*, is not grammatical.

## 2 Excerpting and modelling

In this section, we will focus on the excerpting and further handling of the data, as well as describing the process of statistical modelling.

We worked with the Czech National Corpus (Křen et al., 2015), which has approximately 120 million tokens. Firstly, all the N-BY-N and NUM-BY-NUM occurrences were extracted via CQL and regex queries<sup>1</sup> which resulted in 2 264 and 537 hits, respectively. An example of two such occurrences is in (3). The occurrences were randomized and then a manageable number of them was selected to be annotated later. Overall, we were working with 300 NUM-BY-NUM and 600 N-BY-N constructions (the unequal number of selected constructions illustrates the unequal number of natural occurrences).

- (3) a. Ubíral se blok za blokem dál.  
walked-on.3SG REFL block.MASC.A after block.MASC.INST farther  
'He walked on block by block.' N-BY-N
- b. Němci zapalovali bloky jeden za druhým.  
Germans lit.on.fire.PL blocks.MASC.A one.MASC.N after second.MASC.INST  
'The Germans lit the blocks on fire one by one.' NUM-BY-NUM

<sup>1</sup>Search for N-BY-N: 1: [tag="N.\*"] [lemma="za"] 2: [tag="N.\*"] & 1.lemma=2.lemma; and for NUM-BY-NUM: [lemma="jeden"] [lemma="za"] [lemma="druhý"].

The annotation was carried out manually and each sentence was judged by 4 criteria listed below, all with their respective factor names used later in the statistical model.

1. ANTECEDENT: Is there a local antecedent for the adverbial, such as *bloky* “blocks” for *jeden za druhým* “one by one” in (3-b)? (the mereological approaches decompose the local argument)
2. PREDICATETYPE: Is the predicate scalar? (scalar approaches work with the incremental degree change,  $\delta$ )
3. ACCOMPLISHMENTSTATUS: Is the main verb an accomplishment? (Braginsky and Rothstein (2008) claim that the N-BY-N/NUM-BY-NUM modification is eligible only with accomplishments)
4. PLURACTIONALITYSTATUS: Is the main verb pluractional?

After the tagging, we fitted a logistic regression model<sup>2</sup> and a generalized linear model<sup>3</sup> to recognize which of the 4 factors that differentiate between the two constructions are really distinctive.

Our hypothesis was that speakers choose between using either the N-BY-N or the NUM-BY-NUM adverbial on the basis of their conceptualization of the given event. If they view the sequence of sub-events from the mereological perspective, they tend to use the NUM-BY-NUM construction. On the other hand, the N-BY-N construction is chosen for scalar perspectives.

In both logistic models, the response variable was CONSTRUCTION and the 4 criteria were the predictors (of factor type). The `lrm()` model reported outstanding discrimination: the concordance index  $C = 0.902$ . Table 1 displays the log odds of coefficients, standard errors,  $p$ -values and the exponentiated values of the coefficients for each predictor. The coefficients and the  $p$ -values show that only two conditions, ANTECEDENT and PREDICATETYPE, were statistically significant. The two other conditions, ACCOMPLISHMENTSTATUS and PLURACTIONALITYSTATUS do not show significant differentiation between the N-BY-N and NUM-BY-NUM construction. This is also confirmed by the 95 % confidence interval visualised in Table 2.<sup>45</sup>

Conditions	Coef	S.E.	$p$ -value	Exp. Coefs
INTERCEPT	-1.49	0.65	0.02	0.22
ANTECEDENT=no-antecedent	3.63	0.24	< 0.0001	37.86
PREDICATETYPE=scalar	3.60	0.28	< 0.0001	36.92
ACCOMPLISHMENTSTATUS=not-accomplishment	-0.31	0.65	0.63	0.73
PLURACTIONALITYSTATUS=pluractional	-0.28	0.40	0.48	0.75

Table 1: Statistical results of the linear regression model

The two distinguishing predictors, ANTECEDENT and PREDICATETYPE, show very similar strength. Moreover, the odds of the N-BY-N vs. NUM-BY-NUM construction (the latter was the reference level) in sentences without an antecedent are approximately 38, and in sentences with scalar verbs are approximately 37, thus the odds are similar as well. This shows us that in sentences with clear antecedent and non-scalar verb, speakers strongly tend to select the NUM-BY-NUM construction, whereas in sentences with a scalar verb and without antecedent they strongly prefer the N-BY-N adverbial. The two other factors are statistically non-significant.

<sup>2</sup>We used the `lrm()` function from the R add-on package `rms` (Harrell Jr., 2020) and the following formula: `m.lrm <- lrm(Construction ~ Antecedent + PredicateType + AccomplishmentStatus + PluractionalityStatus, data = NzAn)`

<sup>3</sup>We used the `glm()` function from the R base package (R Core Team, 2013) and the following formula: `m.glm <- glm(Construction ~ Antecedent + PredicateType + AccomplishmentStatus + PluractionalityStatus, data = NzAn, family = binomial)`

<sup>4</sup>The values here are retrieved by the `confint(m.glm)` function. `m.glm()` resulted in similar coefficients and other values as `m.lrm`.

<sup>5</sup>In both Table 1 and Table 2, we report the intercept values which (as usually both in linear and logistic models) represent the expected value of the response variable CONSTRUCTION when all the predictors are at their reference levels.

Conditions	2.5 %	97.5 %
INTERCEPT	-2.75	-0.21
ANTECEDENT=no-antecedent	3.18	4.12
PREDICATE TYPE=scalar	3.07	4.19
ACCOMPLISHMENT STATUS=not-accomplishment	-1.60	0.93
PLURACTIONALITY STATUS=pluractional	-1.05	0.49

Table 2: 95 % confidence interval of the generalized linear model

### 3 Analysis and discussion

In this section, we turn to the formal semantic analysis of the constructions in question, which was modified on the basis of the statistical modelling of the corpus data. We will introduce the essential pieces of instruments which can deal with the new Slavic data and by hypothesis can scale-up to analogical types of constructions and other natural languages, too.

First, we claim, unlike previous authors, that speakers of natural language differentiate between the two constructions and conceptualize linearly ordered pluralities of events either as mereological (resulting in NUM-BY-NUM preference) or as scalar (preferring the N-BY-N construction).

For mereological semantics we follow Brasoveanu and Henderson (2009) and for the NUM-BY-NUM construction in example (2-b), we propose the semantics in (4-a). The contribution of the adverbial is highlighted with the  $\lceil \dots \rceil$  notation and it essentially requires: i) the sub-events  $e'$  (of the plural event  $e$ ) to be temporally sequenced; ii) each sub-event ( $lay$ ) to have its own patient and there to be a plurality of them; iii) each sub-event to be individuated by a particular brick.

For the scalar semantics we assume basically Henderson's (2013) scalar analysis, but we add to it an incrementality requirement formalized in (4-c). (4-b) exemplifies a scalar semantics for the N-BY-N construction, demonstrated here with (2-a), and requires: i) a plurality of events; ii) a linear order; iii) for each increase in the difference function over the  $\theta$ -role ( $\text{th}_\delta$ ) to measure exactly one brick; iv) each sub-event to be smaller than the main event; v) each event  $e''$  in the time following any event  $e'$ , to properly contain the event  $e'$  (indicated by **incr** in (4-c)).

The incrementality addition is our contribution, and we intend it as a fix for too weak truth conditions of Henderson's (2013) scalar semantics, because for Henderson, the incrementality of events is not mapped from the difference function over degrees. That means that (2-a) would be true, for example, if the house was built brick after brick without creating the whole structure. The incrementality is partially inspired by Braginsky and Rothstein (2008), but it is weaker in order to avoid their accomplishment incrementality prediction, which turns out to be empirically wrong.

- (4) a.  $\exists e_\epsilon(\text{LIE}(e) \wedge$   
 $\text{pat}(e) = \sigma x.*\text{BRICKS}(x) \wedge$   
 $\lceil \text{linear.order}(\{e' \leq e : \text{atom}(e')\}) \wedge$   
 $|\{\text{pat}(e') : e' \leq e \wedge \text{atom}(e')\}| > 1 \wedge$   
 $\forall e' \leq e(\text{atom}(e') \rightarrow \text{BRICK}(\text{pat}(e')^\lceil))$
- b.  $\exists e[\text{BUILD}(e) \wedge$   
 $\text{ag}(e) = x \wedge$   
 $\text{th}(e) = \sigma x.*\text{HOUSE}(y) \wedge$   
 $\lceil |\text{atoms}(e)| > 1 \wedge$   
 $\text{linear.order}(\text{atoms}(e)) \wedge$   
 $\forall e' \sqsubseteq e[\text{atom}(e') \rightarrow \text{BRICK}(\text{BUILD}_\delta^{\text{th}}(e')) = \mathbf{1}] \wedge$   
 $\forall e' \sqsubseteq e[\text{BRICK}(\text{BUILD}_\delta^{\text{th}}(e')) < \text{BRICK}(\text{BUILD}_\delta^{\text{th}}(e))] \wedge$   
 $\text{incr}(\text{BUILD}, e)^\lceil]$
- c.  $\text{incr}(P, e) \leftrightarrow \forall e', e'' \sqsubseteq e[(P(e') \wedge P(e'') \wedge \tau(e') \prec \tau(e'')) \rightarrow e' \sqsubseteq e'']$

## 4 Summary

The aim of this paper was to show that native speakers prefer the NUM-BY-NUM construction when conceptualizing the sequence of sub-events as divided into the parts corresponding to the relevant objects, and that they choose the N-BY-N construction if they conceptualize the main event as a plurality of sub-events developing along a scalar dimension. Theoretically speaking, we propose that both mereological and scalar approaches to NUM-BY-NUM and N-BY-N are correct, but each for a different type of data. In Czech, the difference between the theories maps to the two discussed constructions. In the future, we want to investigate whether this approach can scale up to a wider range of data (such as reciprocal constructions or reduplicative event modifiers like *time after time*) discussed by Beck & von Stechow (2007). Nevertheless, the mereological approach of Beck & von Stechow (2007) yields the wrong truth conditions if applied to degree achievements and verbs of motion, as shown convincingly by Henderson (2013).

Still, the linguistic work on these adverbials is far from concluded, since several other aspects remain to be clarified. It would be worth examining, for example, what is the relationship between the alternating prepositions and whether they differ somehow in their semantic contribution to the construction. Future research could also try to answer the question whether the proposed analyses can scale up to a wider variety of cross-linguistic data.

## References

- Sigrid Beck and Arnim von Stechow. 2007. Pluractional adverbials. *Journal of Semantics*, 24(3):215–254.
- Pavel Braginsky and Susan Rothstein. 2008. Vendlerian classes and the Russian aspectual system. *Journal of Slavic linguistics*, pages 3–55.
- Adrian Brasoveanu and Robert Henderson. 2009. Varieties of distributivity: ‘One by One’ vs ‘Each’. In *Semantics and Linguistic Theory*, volume 19, pages 55–72.
- Frank E. Harrell Jr., 2020. *rms: Regression Modeling Strategies*. R package version 6.0-1. Available from <https://CRAN.R-project.org/package=rms>.
- Robert Henderson. 2013. Quantizing scalar change. In *Semantics and Linguistic Theory*, volume 23, pages 473–492.
- M. Křen, V. Cvrček, T. Čapka, A. Čermáková, M. Hnátková, L. Chlumská, T. Jelínek, D. Kovářiková, V. Petkevič, P. Procházka, H. Skoumalová, M. Škrabal, P. Truneček, P. Vondříčka, and A. Zasina. 2015. SYN2015: Reprezentativní korpus psané češtiny. Praha: Ústav Českého národního korpusu, FF UK.
- R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. Available from <https://www.R-project.org/>.